

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-105193

(43)Date of publication of application : 24.04.1998

(51)Int.Cl. G10L 9/00

G10L 9/14

G10L 9/18

H03M 7/30

(21)Application number : 08-254967 (71)Applicant : YAMAHA CORP

(22)Date of filing : 26.09.1996 (72)Inventor : YAMAUCHI KENICHI

(54) SPEECH ENCODING TRANSMISSION SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To listen to speech information reception side without any wait tinge even through a low-speed transmission line and to listen to speech information of high quality in a short wait time even when the same speech information is listened again to with high quality.

SOLUTION: A transmission part divides the encoding output obtained by encoding a speech input signal by a scalable encoder I into outline data of a low bit rate that C2111 be transmitted in real time and detailed data for reproducing the speech signal with high quality by being combined with the outline data. A data transmission part 5 transmits the outline data together prior to the detailed data and then transmits the detailed data together thereafter. A reception part sequentially decodes the received outline data without waiting for the detailed data to be received and reproduces the speech signal in real time. Consequently, light listening becomes possible. The reception part once the outline data and detailed data are both transmitted puts the both together to reproduce the

speech signal with high quality.

LEGAL STATUS [Date of request for examination] 06.02.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3622365

[Date of registration] 03.12.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

*** NOTICES ***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not

reflect the original precisely.

2. **** shows the word which can not be translated.

3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] In the voice coding transmission system which reproduces said sound signal from encoding a sound signal in the transmitting section, transmitting through a transmission line, and receiving and decoding in a receive section said transmitting section For the coding output which encoded and acquired said sound signal, said sound signal The serially refreshable outline data of a low bit rate, Said sound signal is divided into the detail data for reproducing for high quality combining this outline data. It is what transmits said outline data collectively in advance of said detail data, and transmits said detail data collectively after that. Said receive section The voice coding transmission system characterized by being what compounds both and reproduces said sound signal for high quality if sequential decoding of said received outline data is carried out without waiting for reception of said detail data, said sound signal is reproduced serially and both said outline data and detail data are transmitted.

[Claim 2] Said transmitting section is a voice coding transmission system

according to claim 1 characterized by being what controls the bit rate of said outline data according to the situation of said transmission line.

[Claim 3] m bits ($m < n$) of high orders of the n bits (n is two or more integers) PCM (Pulse Code Modulation) data obtained by said transmitting section carrying out analog-to-digital conversion of said sound signal Said outline data, It is what uses a low order $n-m$ bit as said detail data. Said receive section The voice coding transmission system according to claim 1 or 2 characterized by being what reproduces said sound signal serially from m bits of high orders of said PCM data, and reproduces said sound signal for high quality from m bits of high orders of said PCM data, and the complex data of a low order $n-m$ bit.

[Claim 4] It is the voice coding transmission system according to claim 1 or 2 characterized by being what said receive section reproduces said sound signal serially from said monophonic signal, and reproduces stereo voice from said monophonic signal and a differential signal by said transmitting section using the differential signal of the sound signal of said outline data and said multiple channel as said detail data for the monophonic sound signal adding the sound signal of the multiple channel for obtaining stereo voice.

[Claim 5] It is the voice coding transmission system according to claim 1 or 2 characterized by being what said receive section reproduces said sound signal serially from the low-pass component of said sound signal, and reproduces said

sound signal in high quality from the low-pass component and high-frequency component of said sound signal by said transmitting section using the high-frequency component of said outline data and said sound signal as said detail data for the low-pass component of a sound signal.

[Claim 6] Said transmitting section has the 1st and 2nd sign books, and optimal vector selection is performed from the 1st sign book to the target vector based on said sound signal. Next, two-step vector quantization processing which chooses from the 2nd sign book the subvector which approaches a target vector most combining the vector is performed. The index of the vector chosen from said 1st sign book Said outline data, It is what uses as said detail data the index of the vector chosen from said 2nd sign book. Said receive section Said sound signal is serially reproduced from the index of the vector chosen from said 1st sign book. The voice coding transmission system according to claim 1 or 2 characterized by being what reproduces said sound signal in high quality from the index of the vector chosen from said 1st and 2nd sign books.

[Claim 7] Said transmitting section is claim 1 characterized by being what waits for the transmission demand of the detail data sent from said receive section, and transmits said detail data to said receive section thru/or the voice coding transmission system of six given in any 1 term.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to a suitable voice coding transmission system to perform especially real-time transmission of speech information about the voice coding transmission system which encodes and transmits a sound signal.

[0002]

[Description of the Prior Art] When transmitting speech information through a communication line, unless all speech information is transmitted conventionally, playback of speech information is not performed. For this reason, in transmitting speech information to real time and reproducing, it becomes important to choose the optimal transmission bit rate, i.e., coding compressibility, in consideration of a communication band. For example, when it desires a certain amount of playback quality, coding compressibility can seldom be raised, but by the time all speech information is transmitted, the latency time will occur. What is necessary is to raise coding compressibility and just to reduce playback quality to demand of wanting to try listening lightly, even if it lowers some playback quality.

[0003] Conventionally, the scalable voice coding method which can choose the playback quality and coding compressibility at the time of decode as arbitration is proposed for such the object (for example, "coding of musical sound and voice with a scalable layered structure" God Akio, Miki **: the Acoustical Society of Japan lecture collected works, H7.9 voice B3-1-5, pp 277-278). Using three hierarchies' coding method, on the 1st hierarchy, this method encodes the narrow-band signal which carried out the down sampling of the input signal, takes the difference of the playback sound from the input signal which extended the band rather than the lower layer to a lower layer, and encodes this by other high order hierarchies. such a layered structure -- the case (when there are no allowances in a communication band) of a low bit rate -- the information only on a lower layer -- using -- coding of low quality -- performing -- the case (when allowances are in a communication band) of a high bit rate -- the information on the upper layer -- using -- a playback sound -- a broadband -- and it is quality and reproduces.

[0004]

[Problem(s) to be Solved by the Invention] However, in the conventional voice coding transmission system mentioned above, as a result of trying listening speech information lightly on real time by the receiving side, in desiring transmission of more quality speech information, it is necessary to retransmit the

bit stream of a high bit rate from the beginning, and there is a problem that the latency time in that case becomes long, again.

[0005] This invention was made in view of such a trouble, even if it is a low speed transmission line, it can try listening speech information without the latency time by the receiving side, and it aims at offering the voice coding transmission system which can hear the speech information of high quality by little latency time to hear the same speech information again for high quality as a result of an audition.

[0006]

[Means for Solving the Problem] In the voice coding transmission system which reproduces said sound signal from the voice coding transmission system concerning this invention encoding a sound signal in the transmitting section, transmitting it through a transmission line, and receiving and decoding in a receive section Said sound signal for the coding output said whose transmitting section encoded and acquired said sound signal The serially refreshable outline data of a low bit rate, Said sound signal is divided into the detail data for reproducing for high quality combining this outline data. It is what transmits said outline data collectively in advance of said detail data, and transmits said detail data collectively after that. If said receive section does sequential decoding of said received outline data, without waiting for reception of said detail data, and

reproduces said sound signal serially and both said outline data and detail data are transmitted, it will be characterized by being what compounds both and reproduces said sound signal for high quality.

[0007] If the transmitting section transmits the outline data of a refreshable low bit rate serially, since a receive section will reproduce this outline data serially according to this invention, without waiting for reception of detail data, it can try listening the outline of music information without the latency time lightly. Consequently, what is necessary is to wait for transmission of detail data and just to reproduce speech information, when the user of a receiving side wants to hear the still more nearly quality and same speech information. since detail data are compounded with outline data and reproduced, compared with the case where all speech information is again transmitted like before, the latency time is boiled markedly and mitigated. Moreover, when expectation of the user of a receiving side is not answered as a result of the audition of the received outline data, reception of future data can be stopped and useless transmission of the data based on this can be prevented.

[0008] In addition, what is necessary is just to adjust the bit rate of outline data according to the situation of a transmission line, in transmitting speech information using a network where a communication band is changed every moment like the Internet. In this case, what is necessary is just to insert suitably

the information on a data rate, or the information on that change in the bit stream to transmit.

[0009] As outline data, data of m bits of high orders of n-bit PCM data, the addition data of the right-and-left channel in stereo voice, the low-pass component of a sound signal, the vector index of the 1st step in two-step vector quantization, etc. can be used, for example. moreover, the difference of a right-and-left channel [in / corresponding to these as detail data / the data of the low order n-m bit of the above-mentioned PCM data, and the above-mentioned stereo voice] -- data, the high-frequency component of a sound signal, the vector index of the 2nd step in two-step vector quantization, etc. can be used, respectively. thus, since detail data constitute a part of speech information, compared with the total amount of information of the speech information in the case of resending in order to acquire the same quality, it is markedly alike, and the amount of information of itself has it, consequently they can also lessen the latency time. [little]

[0010] Moreover, if the transmitting section waits for the detail data transfer demand sent from a receive section and transmits detail data to a receive section, transmission of useless data can be prevented still more effectively.

[0011]

[Embodiment of the Invention] Hereafter, the gestalt of desirable implementation

of this invention is explained with reference to a drawing. Drawing 1 is the block diagram showing the configuration of the transmitting section in the voice coding transmission system concerning the example of this invention. In addition, as this transmitting section, the method which transmits the outline data of a fixed low-speed bit rate first, and the method which adjusts the bit rate of outline data according to a network situation can be considered regardless of a network situation. Then, the configuration of the transmitting section is hereafter explained about the two above-mentioned kinds of methods, respectively.

[0012] (1) When the bit rate of outline data is immobilization, a voice input signal is divided into the outline data of a fixed bit rate, and the detail data for complementing this outline data and enabling playback of high quality based on the bit rate information directed beforehand while being inputted into the scalable encoder 1 and encoding here based on a predetermined coding method. Outline data are supplied to a multiplexer 2 with additional information, such as a header and a boundary identifier. Once detail data are memorized by the store 3, reading appearance of them is carried out under control of the rearrangement section 4, and they are supplied to a multiplexer 2. A multiplexer 2 carries out sequential supply of the supplied data in the data transmitting section 5 at the sequence of a header, outline data, a boundary identifier, and detail data. The data transmitting section 5 transmits the bit stream supplied

from the multiplexer 2 to a receive section through a transmission line. The bit rate control section 6 determines the bit rate of outline data according to a directions input, and gives bit rate information to the scalable encoder 1. In addition, even if the network which transmits is a low speed, a bit rate is set up by the transmitting side so that it may become the amount of information in which real-time transmission is possible.

[0013] The example of a format of the bit stream assembled by the multiplexer 2 of this transmitting section is shown in drawing 2 (a). A bit stream is assembled in order of a header, outline data, a boundary identifier, and detail data.

[0014] (2) When the network situation Monitoring Department 7 always supervises a network situation through the data transmitting section 5 in this case when the bit rate of outline data is adjustable, and a communication band is changed, give directions information to the bit rate control section 6, and adjust the bit rate of outline data. As the adjustment approach of a bit rate, the coding number of bits can be dropped or approaches, such as thinning out data selectively, can be considered. In this case, the bit rate of outline data and the detail data which make a pair will be created in a form which absorbs change of the bit rate of outline data.

[0015] A format of the bit stream assembled by the multiplexer 2 in case a bit rate is adjustable becomes like drawing 2 (b). That is, the information which

shows that a bit rate changes just before a bit rate changes is inserted in the outline data following a header. Thereby, it becomes possible to recognize the bit rate of outline data in a receive section.

[0016] In addition, detail data are further divided into some, they are rearranged into order with a high significance, reading appearance may be made to be carried out, and you may make it rearrange a frame so that a part with a high significance in music may be transmitted first. As shown in this drawing (b), the part (this example frame numbers 16-20) of the so-called "rust" I want a data origination person to ask to a listener for high quality is arranged to the beginning of detail data, so that it may be shown in the case of music data (a) (for example, drawing 3). Thereby, even when download of detail data is stopped on the way, about the part I want an implementer to hear, it can be heard to a quality playback sound. Rearrangement of such detail data is performed in the rearrangement section 4.

[0017] Drawing 4 is the block diagram showing the configuration of a receive section. Once it is received in the data receive section 11 and the bit stream transmitted through the network is stored in a store 12, it is serially supplied to a decoder 13. A decoder 13 recognizes data after a header is detected until a boundary identifier is detected to be outline data, and starts decoding serially. Although decoding on real time is realized and playback quality falls off a little by

this, sufficient playback sound to try listening lightly is obtained on real time.

[0018] Moreover, even if it lowers a bit rate considerably depending on a network congestion situation, transmission of real time may become impossible. In this case, the network situation Monitoring Department 14 supervises a network situation, and transmits that monitor result to the decoding control section 15. A decoder 15 is controlled serially to start serial decoding with the decoding control section 15 continuous while setting up the latency time like the waiting for 1 minute, 2-minute waiting, and --, corresponding to a network situation and being made not to decode only by receiving data between this latency time serially, after the latency time passes. Thereby, decoding almost near real time without a sound piece is attained.

[0019] Although received following this outline data, detail data enable it to choose whether a listener wishes transmission of the speech information of high quality further, wait to transmit this selection demand to the transmitting section from a receive section, and you may make it transmit detail data by the audition result in a receive section.

[0020] When transmission of detail data is also completed and all detail data are stored in a store 12, the outline data with which the merge section 16 is already transmitted, and the detail data which transmission ended are compounded. When detail data are further divided into plurality, outline data and two or more

detail data will be compounded in the merge section 16. The compounded data are decoded by the decoder 17. Thereby, a sound signal is reproducible for high quality.

[0021] Next, the still more concrete example of the scalable encoder 1 is explained. Drawing 5 is drawing showing the example of a configuration of the scalable encoder 1 with which PCM data were divided into the high order bit and the lower bit, the high order bit was used as outline data, and it used the lower bit as detail data. Analog-to-digital conversion of the voice input signal is carried out with a predetermined sampling frequency with A/D converter 21, for example, it serves as 16-bit PCM data. This PCM data is divided into outline data of 8 bits of high orders, and detail data of 8 bits of low order in the bit division section 22. The additional information generation section 23 generates additional information, such as a header and a boundary identifier, generates and outputs a header at the head of a voice input signal, and generates and outputs a boundary identifier to the end of a voice input signal.

[0022] When a bit rate is fixed, as shown in drawing 6 (a), a bit stream. It is arranged in order of data of 8 bits of low order of a header, data of 8 bits of high orders of PCM data, a boundary identifier, and PCM data, and is transmitted. In a receive section Carry out D / A conversion of the PCM data of 8 bits of high orders, and voice is reproduced on real time. If PCM data of 8 bits of low order

are also transmitted, the voice of high quality will be reproduced by compounding and carrying out D / A conversion of the 16-bit PCM data in corresponding 8 bits of high orders and 8 bits of low order of a frame. Since PCM data of 8 bits of high orders as outline data can reduce a bit rate to 1/2 compared with the case where it transmits in a full bit, voice transmission on real time is attained also in a low-speed network. Moreover, the transmission latency time of PCM data of 8 bits of low order as detail data can also be shortened to one half compared with the case where the data of a full bit are resent.

[0023] In changing a bit rate according to a network situation, according to bit rate information, the bit division section 22 changes the number of bits of a high order, and the low-ranking number of bits, and the additional information generation section 23 outputs the information which shows modification of a bit rate. Supposing PCM data are 16 bits, as shown, for example in drawing 6 (b), it will fix to 8 bits of high orders, and the first outline data following a header will insert the data of "0" (with no bit rate modification), or "1" (a bit rate being subject to change) in the boundary of 8-bit data. And it is 00 as information which shows modification of a bit rate for 2 bits following it when "1" (a bit rate being subject to change) is inserted. -> Increment 01 in 1 bit -> Increment 10 in 2 bit -> 1-bit reduction 11 -> What is necessary is just to make it direct the number of bits to fluctuate like 2-bit reduction.

[0024] Drawing 7 is the example of the scalable encoder 1 when one channel is used as outline data and it uses the remaining channels as detail data among the sound signals of many channels. The 1st and the sound signal of the 2nd channel are stereo signals of for example, a right-and-left channel, and the 3rd channel is for example, a super woofer signal. A/D conversion of the signal of each [these] channel is carried out with A/D converters 31, 32, and 33, respectively. The 1st and the digital data of the 2nd channel are added with an adder 34, and serve as a monophonic signal, and this monophonic signal is outputted as outline data. On the other hand, the 1st and the digital data of the 2nd channel serve as a differential signal (1ch-2ch) with a subtractor 35. And this differential signal and the signal of the 3rd channel are outputted as detail data. The additional information generation section 36 carries out the generation output of the header at the head of a monophonic signal, and carries out the generation output of the boundary identifier at the tail. Thereby, the bit stream generated serves as a header, data of (1ch+2ch), a boundary identifier, and (1ch-2ch) data that continue in order of the data of 3ch(es) like drawing 8.

[0025] Drawing 9 is the block diagram of the scalable encoder 1 at the time of applying this invention to MPEG(Moving Picture Experts Group)-Audio. Input voice is divided into the frequency band of 32 in the subband division section 41, and subsampling is carried out for every band, and it serves as data of 1 band 12

sample extent. In the scaling section 42, it separates into the wave and scale factor which were normalized, and this data is supplied to the quantization section 43. On the other hand, in the FFT (Fast Fourier Transform) section 44, the fast Fourier transform of the input signal is carried out. Based on the called-for voice spectrum, and the wave for every band and scale factor which were called for in the scaling section 42, bit assignment according to the acoustic-sense mental model 45 which also considered the masking effect is performed in the bit quota section 46. According to this bit assignment, the quantization section 43 quantizes the wave for every band, and a scale factor. Among these, the bit stream configuration section 47 outputs the data for a low-pass area part as outline data, and outputs the data for a high-pass area part as detail data. Based on bit rate information, the bit quota section 46 and the bit stream configuration section 47 determine what bit rate is assigned to a part for a low-pass area part. A format of the generated bit stream becomes a header, the data for a low-pass area part, a boundary identifier, and the thing with which the data for a high-pass area part were located in a line in order like drawing 10 (a). Moreover, when a bit rate is made adjustable, as shown in this drawing (b), the information which shows change of a bit rate will be inserted in a part for the direct anterior part of change of the bit rate of the data for the low-pass area part following a header.

[0026] Drawing 11 is the block diagram showing the configuration of the scalable encoder 1 at the time of applying this invention to multistage vector quantization. The data which carried out orthogonal transformation of a voice input signal or this by MDCT (Modified Discrete Cosine Transform) etc. are made into a target vector. In the 1st-step output vector decision section 51, optimal vector selection is performed from the 1st sign book 52 to this target vector. In the 2nd-step output vector decision section 53 The subvector which approaches a target vector most is chosen from the 2nd sign book 54 based on the distorted count result in the distorted count section 55 in combination with the vector chosen in the preceding paragraph. The Maine index chosen from the 1st sign book 52 and the subindex chosen from the 2nd sign book 54 are compounded based on bit rate information in the bit stream configuration section 56, and a bit stream is generated. In this case, since playback of the voice of a certain amount of [the Maine index] quality is possible, outline data and a subindex can be used for the Maine index as detail data.

[0027] When a bit rate is set up fixed, a bit stream is transmitted in order of the subindex as header information, the Maine index as outline data, a boundary identifier, and detail data, as shown in drawing 12 . What is necessary is to operate the Maine index on a curtailed schedule suitably, or to also add a subindex to the Maine index and just to consider as outline data, in changing a

bit rate according to a network situation.

[0028]

[Effect of the Invention] According to this invention, as stated above, the transmitting section transmits the outline data of a refreshable low bit rate serially, and since a receive section is reincarnated serially, without waiting for reception of detail data, this outline data When it can try listening the outline of music information without the latency time lightly and the user of a receiving side moreover wants to hear the still more nearly quality and same speech information That what is necessary is to wait for transmission of detail data and just to reproduce speech information, compared with the former, also boil the latency time in this case markedly, and it is mitigated. And when expectation of the user of a receiving side is not answered as a result of the audition of the received outline data, reception of future data can be stopped and the effectiveness that useless transmission of the data based on this can be prevented is done so.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram of the transmitting section of the voice coding transmission system concerning one example of this invention.

[Drawing 2] It is drawing showing a format of the bit stream transmitted from this transmitting section.

[Drawing 3] It is drawing showing the example of rearrangement of the detail data in this bit stream.

[Drawing 4] It is the block diagram of the receive section of a same sound voice coding transmission system.

[Drawing 5] It is the block diagram showing the 1st example of the scalable encoder in this transmitting section.

[Drawing 6] It is drawing showing a format of the bit stream generated by this encoder.

[Drawing 7] It is the block diagram showing the 2nd example of the scalable encoder in this transmitting section.

[Drawing 8] It is drawing showing a format of the bit stream generated by this encoder.

[Drawing 9] It is the block diagram showing the 3rd example of the scalable encoder in this transmitting section.

[Drawing 10] It is drawing showing a format of the bit stream generated by this encoder.

[Drawing 11] It is the block diagram showing the 4th example of the scalable encoder in this transmitting section.

[Drawing 12] It is drawing showing a format of the bit stream generated by this encoder.

[Description of Notations]

1 [-- The rearrangement section, 5 / -- The data transmitting section, 6 / -- 7 A bit rate control section, 14 / -- The network situation Monitoring Department, 11 / -- A data receive section, 13 / -- It is a decoder and 15 serially. / -- A decoding control section, 16 / -- The merge section, 17 / -- Decoder.] -- A scalable encoder, 2 -- 3 A multiplexer, 12 -- A store, 4

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-105193

(43) 公開日 平成10年(1998) 4月24日

(51) Int.Cl. ⁶	識別記号	F I		
G 1 0 L	9/00	G 1 0 L	9/00	N
	9/14		9/14	J
				N
	9/18		9/18	A
H 0 3 M	7/30	H 0 3 M	7/30	B
審査請求 未請求 請求項の数7 OL (全 7 頁)				

(21) 出願番号 特願平8-254967

(22) 出願日 平成8年(1996) 9月26日

(71) 出願人 000004075

ヤマハ株式会社

静岡県浜松市中沢町10番1号

(72) 発明者 山内 健一

静岡県浜松市中沢町10番1号 ヤマハ株式会社内

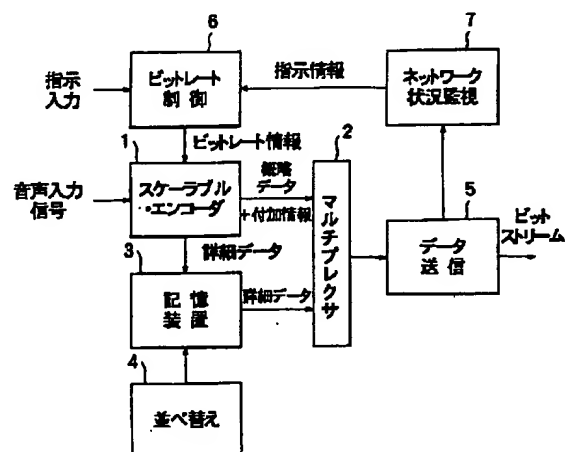
(74) 代理人 弁理士 伊丹 勝

(54) 【発明の名称】 音声符号化伝送方式

(57) 【要約】

【課題】 低速な伝送路であっても受信側で待ち時間なしに音声情報を試聴することができ、試聴の結果、同じ音声情報を高品質で再度聴きたい場合にも、少ない待ち時間で高品質の音声情報を聴くことを可能にする。

【解決手段】 送信部は、音声入力信号をスケーラブル・エンコーダ1で符号化して得た符号化出力を、リアルタイム伝送可能な低ビットレートの概略データと、この概略データと組み合わせて音声信号を高品質で再生するための詳細データとに分割する。データ送信部5は、概略データを詳細データに先立ってまとめて伝送し、その後、詳細データをまとめて伝送する。受信部は、受信された概略データを詳細データの受信を待つことなく逐次復号して音声信号をリアルタイムで再生する。これにより、軽い試聴が可能になる。受信部は、概略データ及び詳細データが共に伝送されたら、両者を合成して音声信号を高品質で再生することができる。



【特許請求の範囲】

【請求項1】 音声信号を送信部で符号化して伝送路を介して伝送し受信部で受信して復号することにより前記音声信号を再生する音声符号化伝送方式において、前記送信部は、前記音声信号を符号化して得た符号化出力を、前記音声信号を逐次再生可能な低ビットレートの概略データと、この概略データと組み合わせて前記音声信号を高品質で再生するための詳細データとに分割し、前記概略データを前記詳細データに先立ってまとめて伝送し、その後前記詳細データをまとめて伝送するものであり、

前記受信部は、受信された前記概略データを前記詳細データの受信を待つことなく逐次復号して前記音声信号を逐次再生し、前記概略データ及び詳細データが共に伝送されたら、両者を合成して前記音声信号を高品質で再生するものであることを特徴とする音声符号化伝送方式。

【請求項2】 前記送信部は、前記伝送路の状況に応じて前記概略データのビットレートを制御するものであることを特徴とする請求項1記載の音声符号化伝送方式。

【請求項3】 前記送信部は、前記音声信号をアナログ・デジタル変換して得られた n ビット(n は2以上の整数)のPCM(Pulse Code Modulation)データの上位 m ビット($m < n$)を前記概略データ、下位 $n-m$ ビットを前記詳細データとするものであり、前記受信部は、前記PCMデータの上位 m ビットから前記音声信号を逐次再生し、前記PCMデータの上位 m ビットと下位 $n-m$ ビットの合成データから前記音声信号を高品質で再生するものであることを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項4】 前記送信部は、ステレオ音声を得るための複数チャネルの音声信号を加算したモノラル音声信号を前記概略データ、前記複数チャネルの音声信号の差分信号を前記詳細データとするものであり、前記受信部は、前記モノラル信号から前記音声信号を逐次再生し、前記モノラル信号及び差分信号からステレオ音声を再生するものであることを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項5】 前記送信部は、音声信号の低域成分を前記概略データ、前記音声信号の高域成分を前記詳細データとするものであり、

前記受信部は、前記音声信号の低域成分から前記音声信号を逐次再生し、前記音声信号の低域成分及び高域成分から前記音声信号を高品質に再生するものであることを特徴とする請求項1又は2記載の音声符号化伝送方式。

【請求項6】 前記送信部は、第1及び第2の符号帳を有し、前記音声信号に基づくターゲットベクトルに対して第1の符号帳から最適なベクトル選択を行い、次にそのベクトルと組み合わせて最もターゲットベクトルに近づくサブベクトルを第2の符号帳から選択する2段階ベクトル量子化処理を実行し、前記第1の符号帳から選択さ

れたベクトルのインデックスを前記概略データ、前記第2の符号帳から選択されたベクトルのインデックスを前記詳細データとするものであり、

前記受信部は、前記第1の符号帳から選択されたベクトルのインデックスから前記音声信号を逐次再生し、前記第1及び第2の符号帳から選択されたベクトルのインデックスから前記音声信号を高品質に再生するものであることを特徴とする請求項1又は2記載の音声符号化伝送方式。

10 【請求項7】 前記送信部は、前記受信部から送られる詳細データの伝送要求を待って前記詳細データを前記受信部に伝送するものであることを特徴とする請求項1乃至6のいずれか1項記載の音声符号化伝送方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声信号を符号化して伝送する音声符号化伝送方式に関し、特に音声情報のリアルタイム伝送を行うのに好適な音声符号化伝送方式に関する。

20 【0002】

【従来の技術】音声情報を通信回線を介して伝送する場合、従来は、全ての音声情報が伝送されないと音声情報の再生が行われない。このため、音声情報をリアルタイムに伝送して再生する場合には、通信帯域を考慮して最適な伝送ビットレート、即ち符号化圧縮率を選択することが重要となる。例えば、ある程度の再生品質を望む場合、符号化圧縮率をあまり高めることができず、音声情報が全て伝送されるまでに待ち時間が発生する。再生品質を多少落としても、軽く試聴してみたいという要求に対しては、符号化圧縮率を高めて再生品質を低下させればよい。

30 【0003】従来、このような目的で、復号時の再生品質や符号化圧縮率を任意に選択できるスケーラブルな音声符号化方式が提案されている(例えば「スケーラブルな階層構造を持つ楽音・音声の符号化」神明夫、三樹聡：日本音響学会講演論文集，H7.9音声B3-1-5，pp277-278)。この方式は、3階層の符号化方式を用い、第1階層では入力信号をダウンサンプリングした狭帯域信号を符号化し、その他の上位階層では、下位層よりも帯域を上げた入力信号から下位層までの再生音の差分をとり、これを符号化する。このような階層構造によって、低ビットレートの場合(通信帯域に余裕がない場合)は、下位層のみの情報を用いて低品質の符号化を実行し、高ビットレートの場合(通信帯域に余裕がある場合)は、上位層の情報も用いて、再生音を広帯域且つ高品質で再生する。

40 【0004】

【発明が解決しようとする課題】しかしながら、上述した従来の音声符号化伝送方式では、受信側で音声情報をリアルタイムで軽く試聴した結果、より品質の良い音声

情報の伝送を望む場合には、再度、高ビットレートのビットストリームを最初から伝送し直す必要があり、その際の待ち時間が長くなるという問題がある。

【0005】この発明は、このような問題点に鑑みなされたもので、低速な伝送路であっても受信側で待ち時間なしに音声情報を試聴することができ、試聴の結果、同じ音声情報を高品質で再度聴きたい場合にも、少ない待ち時間で高品質の音声情報を聴くことが出来る音声符号化伝送方式を提供することを目的とする。

【0006】

【課題を解決するための手段】この発明に係る音声符号化伝送方式は、音声信号を送信部で符号化して伝送路を介して伝送し受信部で受信して復号することより前記音声信号を再生する音声符号化伝送方式において、前記送信部が、前記音声信号を符号化して得た符号化出力を、前記音声信号を逐次再生可能な低ビットレートの概略データと、この概略データと組み合わせて前記音声信号を高品質で再生するための詳細データとに分割し、前記概略データを前記詳細データに先立ってまとめて伝送し、その後前記詳細データをまとめて伝送するものであり、前記受信部が、受信された前記概略データを前記詳細データの受信を待つことなく逐次復号して前記音声信号を逐次再生し、前記概略データ及び詳細データが共に伝送されたら、両者を合成して前記音声信号を高品質で再生するものであることを特徴とする。

【0007】この発明によれば、送信部が逐次再生可能な低ビットレートの概略データを送信すると、受信部がこの概略データを、詳細データの受信を待つことなく逐次再生するので、音楽情報の概要を待ち時間なしに軽く試聴することができる。この結果、受信側のユーザが、更に高品質で同じ音声情報を聴きたい場合には、詳細データの伝送を待つて音声情報の再生を行えばよい。詳細データは、概略データと合成されて再生されるので、従来のように全ての音声情報を再度伝送する場合に比べ、待ち時間は格段に軽減される。また、受信された概略データの試聴の結果、受信側のユーザの期待に添わなかった場合には、以後のデータの受信を中止することができ、これによるデータの無駄な伝送が防止できる。

【0008】なお、インターネットのように通信帯域が刻々と変動するようなネットワークを使用して音声情報を伝送する場合には、伝送路の状況に応じて概略データのビットレートを調整すればよい。この場合、伝送するビットストリームにデータレートの情報又はその変化の情報を適宜挿入しておけばよい。

【0009】概略データとしては、例えばnビットのPCMデータの上位mビットのデータ、ステレオ音声における左右チャンネルの加算データ、音声信号の低域成分、2段ベクトル量子化における第1段のベクトルインデックス等を用いることができる。また、詳細データとしては、これらに対応して例えば上記PCMデータの低位n

mビットのデータ、上記ステレオ音声における左右チャンネルの差分データ、音声信号の高域成分、2段ベクトル量子化における第2段のベクトルインデックス等をそれぞれ用いることができる。このように、詳細データは、音声情報の一部を構成するものであるから、それ自体の情報量は、同一の品質を得るために再送する場合の音声情報の全情報量に比べて格段に少なく、その結果、待ち時間も少なくすることができる。

【0010】また、送信部が受信部から送られる詳細データの転送要求を待つて詳細データを受信部に伝送するものであれば、無駄なデータの伝送を更に効果的に防ぐことができる。

【0011】

【発明の実施の形態】以下、図面を参照して、この発明の好ましい実施の形態について説明する。図1は、この発明の実施例に係る音声符号化伝送方式における送信部の構成を示すブロック図である。なお、この送信部としては、ネットワークの状況の如何に拘わらず、固定的な低速ビットレートの概略データを最初に送信する方式と、ネットワークの状況に応じて概略データのビットレートを調整する方式とが考えられる。そこで、以下、上記2通りの方式について送信部の構成をそれぞれ説明する。

【0012】(1) 概略データのビットレートが固定の場合

音声入力信号は、スケーラブル・エンコーダ1に入力され、ここで所定の符号化方式に基づいて符号化されると共に、予め指示されたビットレート情報に基づいて固定的なビットレートの概略データと、この概略データを補完して高品質の再生を可能にするための詳細データとに分割される。概略データは、ヘッダや境界識別子等の付加情報と共にマルチプレクサ2に供給される。詳細データは、記憶装置3に一旦記憶されたのち、並べ替え部4の制御のもとで読み出され、マルチプレクサ2に供給される。マルチプレクサ2は、供給されたデータをヘッダ、概略データ、境界識別子、詳細データの順番に、データ送信部5に順次供給する。データ送信部5は、マルチプレクサ2から供給されたビットストリームを伝送路を介して受信部に送信する。ビットレート制御部6は、指示入力に従って概略データのビットレートを決定し、ビットレート情報をスケーラブル・エンコーダ1に与える。なお、ビットレートは、伝送を行うネットワークが低速であっても、リアルタイム伝送が可能な情報量となるように送信側で設定する。

【0013】この送信部のマルチプレクサ2で組み立てられるビットストリームのフォーマット例を図2(a)に示す。ビットストリームは、ヘッダ、概略データ、境界識別子、詳細データの順に組み立てられる。

【0014】(2) 概略データのビットレートが可変の場合

この場合、ネットワーク状況監視部7がデータ送信部5を介してネットワークの状況を常に監視し、通信帯域が変動した場合には、ビットレート制御部6に指示情報を与えて概略データのビットレートを調整する。ビットレートの調整方法としては、符号化ビット数を落としたり、部分的にデータを間引く等の方法が考えられる。この場合、概略データと対をなす詳細データのビットレートは、概略データのビットレートの変化を吸収するような形で作成されることになる。

【0015】ビットレートが可変の場合のマルチプレクサ2で組み立てられるビットストリームのフォーマットは、図2(b)のようになる。即ち、ヘッダに続く概略データには、ビットレートの変化する直前にビットレートが変化することを示す情報が挿入される。これにより、受信部で概略データのビットレートを認識することが可能になる。

【0016】なお、詳細データは、更にいくつかに分割され、重要度の高い順に並び替えて読み出されるようにしても良いし、曲の中の重要度の高い部分が最初に送信されるようにフレームを並べ替えるようにしても良い。音楽データの場合、例えば図3(a)に示すように、データ作成者がリスナに高品質で聴いてほしい、いわゆる“さび”の部分(この例ではフレーム番号16~20)を、同図(b)に示すように詳細データの最初に配置する。これにより、詳細データのダウンロードを途中で中止した場合でも、作成者が聴いてほしい部分については、高品質な再生音で聴くことが出来る。このような詳細データの並べ替えは、並べ替え部4にて行われる。

【0017】図4は、受信部の構成を示すブロック図である。ネットワークを介して伝送されたビットストリームは、データ受信部11で受信され、記憶装置12に一旦格納されたのち、逐次デコーダ13に供給される。逐次デコーダ13は、ヘッダが検出されてから境界識別子が検出されるまでのデータを概略データであると認識して、逐次デコードを開始する。これにより、リアルタイムでのデコードが実現され、再生品質は若干落ちるものの、軽く試聴するには十分の再生音がリアルタイムで得られる。

【0018】また、ネットワークの混雑状況によっては、ビットレートをかなり下げてもリアルタイムの伝送が不可能になる場合がある。この場合には、ネットワーク状況監視部14がネットワークの状況を監視して、その監視結果をデコード制御部15に伝送する。デコード制御部15は、ネットワークの状況に応じて、例えば1分待ち、2分待ち、…のように、待ち時間を設定し、この待ち時間の間は、データを受信するだけで逐次デコードを行わないようにすると共に、待ち時間が経過した後は連続的な逐次デコードを開始するように逐次デコーダ15を制御する。これにより、音切れのないほぼリアルタイムに近いデコードが可能になる。

【0019】詳細データは、この概略データに続いて受信されることになるが、受信部での試聴結果によってリスナが更に高品質の音声情報の伝送を希望するか否かを選択できるようにしておき、この選択要求が受信部から送信部に伝送されるのを待って、詳細データを送信するようにしても良い。

【0020】詳細データの送信も完了し、記憶装置12に詳細データが全て格納された場合には、データ合成部16が既に伝送されている概略データと伝送が終了した詳細データとを合成する。詳細データを更に複数に分割した場合には、データ合成部16で概略データと複数の詳細データとが合成されることになる。合成されたデータは、デコーダ17によってデコードされる。これにより、音声信号を高品質で再生することができる。

【0021】次に、スケーラブル・エンコーダ1の更に具体的な実施例について説明する。図5は、PCMデータを上位ビットと下位ビットとに分割し、上位ビットを概略データ、下位ビットを詳細データとしたスケーラブル・エンコーダ1の構成例を示す図である。音声入力信号は、A/D変換器21で所定サンプリング周波数でアナログ・デジタル変換され、例えば16ビットのPCMデータとなる。このPCMデータは、ビット分割部22において、例えば上位8ビットの概略データと、下位8ビットの詳細データとに分割される。付加情報生成部23は、ヘッダや境界識別子等の付加情報を生成するもので、音声入力信号の先頭にヘッダを生成して出力し、音声入力信号の終わりに境界識別子を生成して出力する。

【0022】ビットレートが固定的である場合には、図6(a)に示すように、ビットストリームは、ヘッダ、PCMデータの上位8ビットのデータ、境界識別子、PCMデータの下位8ビットのデータの順に並べられて伝送され、受信部では、上位8ビットのPCMデータをデジタル・アナログ変換してリアルタイムで音声再生し、下位8ビットのPCMデータも伝送されたら、対応するフレームの上位8ビットと下位8ビットとで16ビットのPCMデータを合成してデジタル・アナログ変換することにより、高品質の音声再生する。概略データとしての上位8ビットのPCMデータは、フルビットで送信する場合に比べてビットレートを1/2まで低下させることができるので、低速のネットワークでもリアルタイムでの音声伝送が可能になる。また、詳細データとしての下位8ビットのPCMデータの伝送待ち時間も、フルビットのデータを再送する場合に比べ、1/2に短縮することができる。

【0023】ネットワーク状況に応じてビットレートを変更する場合には、ビットレート情報に応じてビット分割部22が上位のビット数と下位のビット数とを変更し、付加情報生成部23は、ビットレートの変更を示す情報を出力する。PCMデータが16ビットであるとす

10

20

30

40

50

ると、例えば図6(b)に示すように、ヘッダに続く最初の概略データは、上位8ビットに固定しておき、8ビットのデータの境界に“0”(ビットレート変更なし)又は“1”(ビットレートの変更あり)のデータを挿入しておく。そして、“1”(ビットレートの変更あり)が挿入された場合には、それに続く2ビットをビットレートの変更を示す情報として、

00 → 1ビット増加

01 → 2ビット増加

10 → 1ビット減少

11 → 2ビット減少

のように、増減するビット数を指示するようにすればよい。

【0024】図7は、多チャネルの音声信号のうち1チャネルを概略データ、残りのチャネルを詳細データとした場合のスケラブル・エンコーダ1の例である。第1及び第2チャネルの音声信号は、例えば左右チャネルのステレオ信号であり、第3チャネルは、例えばスーパーウーハー信号である。これら各チャネルの信号は、A/D変換器31、32、33でそれぞれA/D変換される。第1及び第2チャネルのデジタルデータは、加算器34で加算されてモノラル信号となり、このモノラル信号が概略データとして出力される。一方、第1及び第2チャネルのデジタルデータは、減算器35で差分信号(1ch-2ch)となる。そして、この差分信号と第3チャネルの信号とが詳細データとして出力される。付加情報生成部36は、モノラル信号の先頭でヘッダを生成出力し、末尾で境界識別子を生成出力する。これにより、生成されるビットストリームは、図8のように、ヘッダ、(1ch+2ch)のデータ、境界識別子、(1ch-2ch)及び3chのデータの順に続くデータとなる。

【0025】図9は、MPEG(Moving Picture Experts Group) -Audioにこの発明を適用した場合のスケラブル・エンコーダ1のブロック図である。入力音声は、サブバンド分割部41で例えば32の周波数帯域に分割され、各バンド毎にサブサンプリングされて1バンド12サンプル程度のデータとなる。このデータがスケラリング部42で、正規化された波形と倍率とに分離され、量子化部43に供給される。一方、入力信号は、FFT(Fast Fourier Transform)部44において、高速フーリエ変換される。求められた音声スペクトラムと、スケラリング部42で求められたバンド毎の波形及び倍率とに基づき、マスキング効果も加味した聴覚心理モデル45に応じたビット割り当てがビット割り当て部46にて行われる。このビット割り当てに応じて量子化部43が各バンド毎の波形と倍率とを量子化する。ビットストリーム構成部47は、このうち、低域部分のデータを概略データとして出力し、高域部分のデータを詳細データとして出力する。低域部分にどの程度のビットレートを割り当てるかは、ビットレート情報に基づき、ビット

割り当て部46及びビットストリーム構成部47が決定する。生成されたビットストリームのフォーマットは、図10(a)のように、ヘッダ、低域部分のデータ、境界識別子、高域部分のデータが順に並んだものとなる。また、ビットレートを可変にした場合には、同図(b)のように、ヘッダに続く低域部分のデータのビットレートの変化の直前部分にビットレートの変化を示す情報が挿入されることになる。

【0026】図11は、多段ベクトル量子化にこの発明を適用した場合のスケラブル・エンコーダ1の構成を示すブロック図である。音声入力信号又はこれをMDC T(Modified Discrete Cosine Transform)等で直交変換したデータをターゲットベクトルとし、第1段出力ベクトル決定部51では、このターゲットベクトルに対して第1の符号帳52から最適なベクトル選択を行い、第2段出力ベクトル決定部53では、前段で選択されたベクトルとの組み合わせで最もターゲットベクトルに近づくサブベクトルを歪計算部55での歪計算結果に基づいて第2の符号帳54から選択する。第1の符号帳52から選択されたメインインデックスと第2の符号帳54から選択されたサブインデックスとがビットストリーム構成部56でビットレート情報に基づいて合成されてビットストリームが生成される。この場合、メインインデックスだけでも、ある程度の品質の音声の再生は可能であるため、メインインデックスを概略データ、サブインデックスを詳細データとして用いることができる。

【0027】ビットレートを固定的に設定した場合には、ビットストリームは、図12に示すように、ヘッダ情報、概略データとしてのメインインデックス、境界識別子、詳細データとしてのサブインデックスの順に伝送される。ビットレートをネットワーク状況に応じて変化させる場合には、メインインデックスを適当に間引きしたり、メインインデックスにサブインデックスも加えて概略データとすればよい。

【0028】

【発明の効果】以上述べたように、この発明によれば、送信部が逐次再生可能な低ビットレートの概略データを送信し、受信部がこの概略データを、詳細データの受信を待つことなく逐次再生するので、音楽情報の概要を待ち時間なしに軽く試聴することができ、しかも受信側のユーザが更に高品質で同じ音声情報を聴きたい場合には、詳細データの伝送を待って音声情報の再生を行えばよく、この場合の待ち時間も従来に比べて格段に軽減され、且つ受信された概略データの試聴の結果、受信側のユーザの期待に添わなかった場合には、以後のデータの受信を中止することができ、これによるデータの無駄な伝送が防止できるという効果を奏する。

【図面の簡単な説明】

【図1】 この発明の一実施例に係る音声符号化伝送方式の送信部のブロック図である。

【図2】 同送信部から送信されるビットストリームのフォーマットを示す図である。

【図3】 同ビットストリームにおける詳細データの並べ替えの例を示す図である。

【図4】 同音声符号化伝送方式の受信部のブロック図である。

【図5】 同送信部におけるスケーラブル・エンコーダの第1の実施例を示すブロック図である。

【図6】 同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

【図7】 同送信部におけるスケーラブル・エンコーダの第2の実施例を示すブロック図である。

【図8】 同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

【図9】 同送信部におけるスケーラブル・エンコーダ*

*の第3の実施例を示すブロック図である。

【図10】 同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

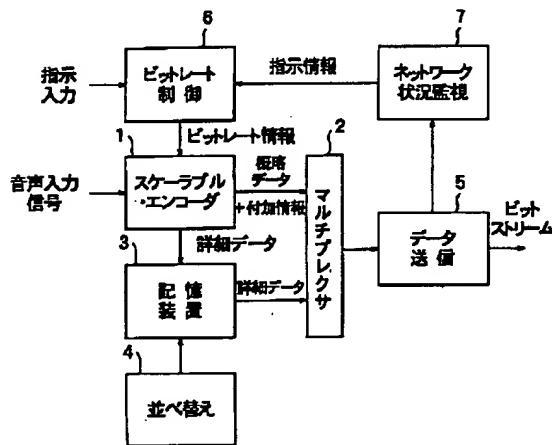
【図11】 同送信部におけるスケーラブル・エンコーダの第4の実施例を示すブロック図である。

【図12】 同エンコーダにより生成されるビットストリームのフォーマットを示す図である。

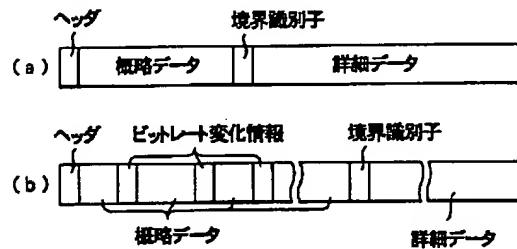
【符号の説明】

1…スケーラブル・エンコーダ、2…マルチプレクサ、3、12…記憶装置、4…並べ替え部、5…データ送信部、6…ビットレート制御部、7、14…ネットワーク状況監視部、11…データ受信部、13…逐次デコーダ、15…デコード制御部、16…データ合成部、17…デコーダ。

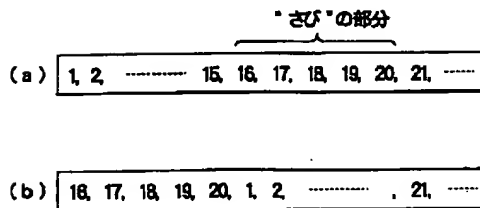
【図1】



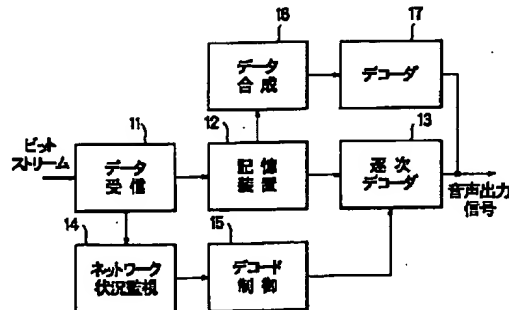
【図2】



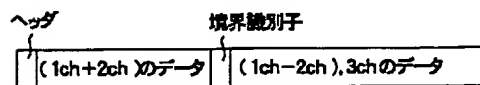
【図3】



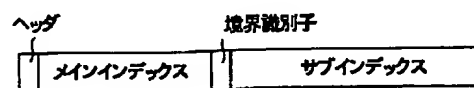
【図4】



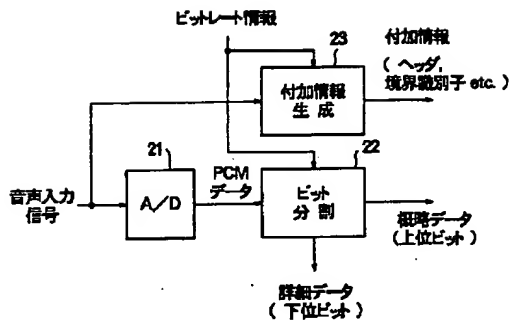
【図8】



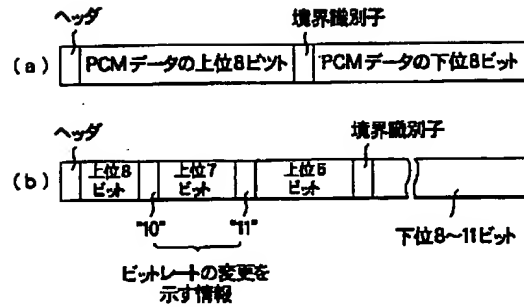
【図12】



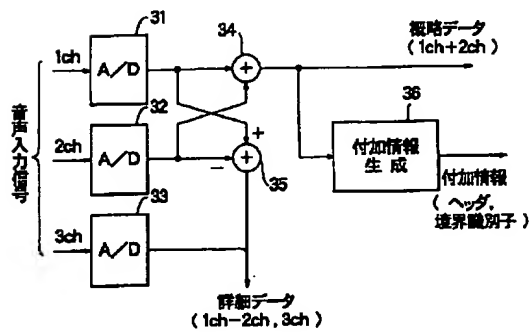
【図5】



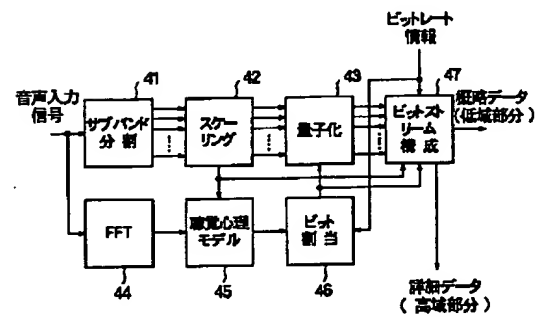
【図6】



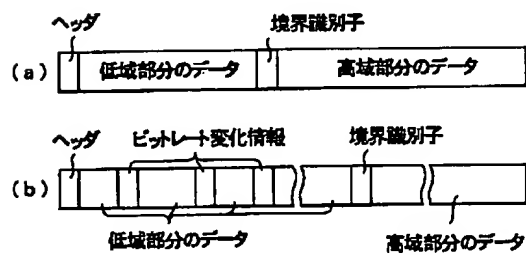
【図7】



【図9】



【図10】



【図11】

